Rule A and Responsibility: A Defense of the Compatibility of Moral Responsibility and Causal Determinism

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation *with research distinction* in Philosophy in the undergraduate colleges of The Ohio State University

By

Benjamin Flowers

The Ohio State University
November, 2008

Project Adviser: Professor Timothy Schroeder, Department of Philosophy

I. Moral Responsibility and Determinism: The General Problem: Moral responsibility is seen by most to be an attribute that we as humans quite obviously possess, and moreover, something on which we place extraordinary value. That being said, our world is increasingly being thought of as a deterministic one. By determined, I mean causally determined, and by this I mean to say that there is exactly one physically possible future at any instant.[1] In other words, given all of the physical laws and the current state of the universe, we could correctly predict each subsequent state of affairs. Like the rest of the population, most philosophers are loathe to give up something as important, seemingly apparent, and commonsensical as moral responsibility, though it is not clear how this could comport with the truth of causal determinism.[2] After all, it seems impossible to get to a state in which we are morally responsible, by means of a mechanism (more specifically, the laws of nature) which dictates the states that we will be in and so the actions which we bring about.

Despite the intuitive force of such an argument, there might be a way for these seemingly disparate and incongruous ideas to sensibly co-exist, and so perhaps a causally determined world containing humans with moral responsibility can exist. In the following essay, I will defend this view, which is known as compatibilism. More specifically, I will defend compatibilism against a very specific and powerful argument: the Direct

[1] Peter van Inwagen, *An Essay on Free Will*, (Oxford: Oxford University Press, 1986), pg. 3.
[2] As Fischer and Ravizza point out [John Martin Fischer and Mark Ravizza, S.J., *Responsibility and Control* (New York: Cambridge University Press, 2000), 15.] the existence of quantum mechanics as a legitimate theory does cast some doubt on the truth of causal determinism. That being said, this paper is concerned (at a very general level) with the compatibility of causal determinism and moral responsibility, *if* the former is true. Moreover, even if it is not but something like quantum mechanics is true, "macroscopic events," as they put it, would be "very close to being determined," and so issues very similar to those involved in the debate between the compatibility of causal determinism and moral responsibility would still arise.

Argument for Incompatibilism, by Peter van Inwagen.[3] After explaining what exactly this argument is, I will examine an important attempted refutation of the Direct Argument, before arguing that said attempt will not work. Finally, I will present a new line of thought which I think provides a more promising route for resisting van Inwagen. In short, I will argue that despite its clear intuitive power, the Direct Argument fails to give us good reason to believe that moral responsibility and causal determinism must be at odds.

II. The Direct Argument for Incompatibilism: Before proceeding, I would like to make a brief but important note. Strictly speaking, the topic at hand will not be compatibilism, which holds that free will, in addition to responsibility, is compatible with determinism. Instead I will be defending what is called semi-compatibilism, or the belief that moral responsibility and causal determinism are compatible, regardless of whether free will is compatible with determinism.[4] This is allowable, since the Direct Argument is meant to challenge just the compatibility of responsibility and determinism, and makes no mention of free will.[5] With that being said, if van Inwagen's Direct Argument is successful, compatibilism and not just semi-compatibilism fails. Because of this, provided that we move using tenets that both views would accept (as I will) we can attempt to clear the way for both views by knocking down the Direct Argument.

---

[3] See: Peter van Inwagen, "The Incompatibility of Responsibility and Determinism," in *Action and Responsibility*, ed. Michael Bradie and Myles Brand (Bowling Green, Ohio: The Applied Philosophy Program at Bowling Green State University, 1980), 30-36. *as well as*: Peter van Inwagen, "What Our Not Having Free Will Would Mean," Chap. 5 in *An Essay on Free Will*.

[4] Mark Ravizza, "Semi-Compatibilism and the Transfer of Non-Responsibility," *Philosophical Studies 75*, (1994): 62.

[5] Peter van Inwagen, "The Incompatibility of Responsibility and Determinism," 30. Van Inwagen would claim that free will is required for moral responsibility, but this is a topic that I will be unable to address in this paper.

The question then, is if causal determinism precludes responsibility,[6] and as has been stated, van Inwagen believes that the answer should be in the affirmative. He arrives at this conclusion by way of the Direct Argument. The argument uses several variables, starting with S, which stands for some sentence that gives a "complete and accurate description of the past."[7] In other words, S is a sentence that tells us all and only the physical truths about some state of the world at a previous point in time. L is a sentence which gives a complete and accurate description of the laws of nature. T represents "any truth whatever,"[8] while □ represents "truth in all possible worlds."[9] Finally, van Inwagen uses the abbreviation Np to mean "p and no human being, or group of human beings, is even partly morally responsible for the fact that p,"[10] where p is a proposition. For instance, if no one is responsible for the truth of the proposition that Denise fell into the pond, this could be represented as: "N Denise fell into the pond." This should be read as "Denise fell into the pond, and no human is even partly morally responsible for the fact that Denise fell into the pond."

Van Inwagen also specifies two inference rules:

(Rule A): $\Box p \vdash Np$: i.e. The fact that it is true in all possible worlds that p, entails that p, and that no human is even partly responsible for the fact that p.

(Transfer NR):[11] $Np, N(p \rightarrow q) \vdash Nq$: i.e. the fact that p and that no human is even partly responsible for p, coupled with the fact that if p then q, and that no human being is

[6] I will use responsibility (as van Inwagen does) interchangeably with "moral responsibility"
[7] Ibid, 31.
[8] Though he says this, I think we can assume van Inwagen to mean "any true state of affairs" whatsoever, as the argument makes no sense if we were to insert, say an *a priori* truth into T.
[9] Ibid, 31.
[10] Ibid, 32.
[11] Peter van Inwagen calls this "rule B," but I will stick with what most are calling in papers about this principle today, and refer to it as Transfer NR.

even partly responsible for this conditional, means that q and that no one is even partly morally responsible for q.

The argument's logical form (included in the footnotes are descriptions of each portion of the argument in non-formal language), is as follows:

(1) □ (S & L. → T))[12]	From the definition of determinism
(2) □ (S→(L→T))[13]	From 1, by logical translation
(3) N(S→(L → T))[14]	From 2 by way of A
(4) NS[15]	Assumption
(5) N(L → T)[16]	From 3 and 4 by way of Transfer NR
(6) NL[17]	Assumption
(7) NT[18]	From 5 and 6 by Transfer NR

The problem for the semi-compatibilist is now clear. Given the truth of determinism, any actual state of affairs (past, present or future) can be substituted for T, since any past state of affairs conjoined with the laws of nature entail that all and only the future states of affairs that do obtain, will obtain. However, we are left with the

[12] It is true in all possible worlds that if some initial state obtains, and the deterministic laws of nature do as well, then T (any state of the world) obtains as well. Van Inwagen rightly sees this as the definition of causal determinism in logical form, since the conjunction of a given state and a given set of laws entail a certain state of affairs (or truth) in the future which we can slide into T. Once we know which state of affairs and laws we are working with, we know all and only which T's will obtain.

[13] It is true in all possible worlds that if the initial state of the world obtains, then if the laws of nature obtain, then T will obtain

[14] Rule A allows us to move from the fact that it is true in all possible worlds that if the initial state of the world obtains, then if the laws of nature obtain, then T will obtain, to the fact that no one is even partly morally responsible for this conditional.

[15] S, a state occurring before any responsible finite agents obtains, and no one is even partly responsible for this fact. This state should obtain before any responsible finite agents, since S must be a state of affairs that no one is even partly morally responsible for, and clearly no one is even partly responsible for a state of affairs occurring before anything with the ability to be responsible existed.

[16] If the laws of nature obtain, then T, and no one is even partly morally responsible for the fact that this conditional is true.

[17] The laws of nature do obtain, and no one is even partly responsible for this fact.

[18] T obtains, and no one is even partly morally responsible for the fact that T obtains.

unpleasant conclusion that T, and no human is even partly morally responsible for T. Given that any state of affairs can be inserted for T, no one is responsible for any state of affairs whatsoever, and so determinism and moral responsibility are incompatible. In summary, if the world is causally determined, there is no state of affairs for which we could be responsible.

What is also clear about this argument is that it depends for its success on the success of two principles: Transfer NR and Rule A. Van Inwagen finds Rule A to be beyond contention, since it seems clear that something which is true in all possible worlds is not the sort of thing that one could be responsible for. I am not so sure about this, and I will address this point later. Regardless, most of the literature on this topic focuses on Transfer NR, which van Inwagen himself admits is contentious. Though he acknowledges his inability to *prove* Transfer NR, he finds it intuitively plausible. Certainly, Transfer NR seems to capture the pre-philosophical intuitions of many. If there was some state p that occurred long before humans existed, such that no one was even partly responsible for p, and if no one is even partly morally responsible for the laws of nature and the fact that they guarantee q given p, then it would seem to follow that we cannot be responsible for q. There might be wiggle room available however, as van Inwagen himself notes:

> But if anyone thinks my belief [in the incompatibilism of responsibility and causal determinism] is false and does for some reason take an interest in my intellectual welfare, here is what he will have to do to get me to see the light: he will have to produce some proposition intuitively more plausible that the proposition that [Transfer NR] is valid and show that this proposition entails [semi-compatibilism], or else he will have to devise a counterexample to [Transfer NR] whose status as such can be established without assuming that determinism and moral responsibility are compatible.[19]

[19] Ibid, 36.

I will look at one attempt at taking the latter route in the next section, before explaining why this method will not help. Though it is my opinion that the Direct Argument might be weakest where van Inwagen thinks it is safe, specifically with regard to Rule A, I will table concerns about this for the moment in order to look more closely at the often discussed Transfer NR, if for no other reason than most prominent attacks on the Direct Argument have been directed at Transfer NR. Further, even if the Direct Argument ends up being a failure on a different count, *something* should be said about this principle, which would seem to independently present a problem to the compatibilist view. With this in mind, I will respond to Transfer NR in Section IV, before moving on to a bigger and more important criticism of Rule A in Section V. Success in these theatres will allow the compatibilist to escape the teeth of van Inwagen's Direct Argument.

III. Transfer NR and Frankfurt Examples: In Chapter 6 of their book *Responsibility and Control*, John Martin Fischer and Mark Ravizza present what they believe to be a type of counterexample to Transfer NR, and in doing so challenge the validity of this argument form.[20] Though there are a slew of other challenges to Transfer NR by a variety of disputants, in my opinion the work by Fischer and Ravizza on the invalidity of Transfer NR is the best to date, and for this reason I will focus my attention on what they have to say. Fischer and Ravizza's arguments rely on what are known as Frankfurt examples, the famous thought experiments named after their inventor, Harry Frankfurt.[21] All follow a basic format, meant to challenge the notion that the ability to have done otherwise is necessary for moral responsibility. This idea that, in order to be

[20] Many of the same arguments can be found in the paper "Semi-Compatibilism and the Transfer of Non-Responsibility," published earlier and independently by Mark Ravizza.
[21] Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy v.66*, no. 23 (1969): 829-839.

responsible for something, one must have been able to have done other than he or she actually did, is known as the Principle of Alternative Possibilities, or hereafter, PAP. This principle is straightforwardly problematic for the compatibilist, since it contradicts the compatibilist idea that, even though in a deterministic universe we would be unable to do otherwise than we actually did (determinism excludes alternate physical possibilities by holding that there is only one physically possible future), we could still be responsible for our actions.[22]

To construct a Frankfurt example, we simply posit a person who desires to do some act, and a "counterfactual intervener" who will prevent that person from doing anything but that act, even should the person attempt to change her mind. The person never changes her mind so the counterfactual intervener does nothing, and though the person could not have done differently (had she tried the intervener would have stepped in), she still seems to be morally responsible.

To make this less abstract, we can imagine an example involving Peter and Harry. Peter wants to meet up with Harry tomorrow in Central Park so that they can steal a bicycle. Harry is really excited about the prospect of thievery, but he is also aware of Peter's fickle nature, and worries that he will change his plans. Because of this Harry, an accomplished neurosurgeon, sneaks into Peter's room in the middle of the night, and implants a chip in his brain which will track Peter's desires. Should Peter change his mind and decide against stealing the bike, Harry will be able to force him to do it anyway, by way of the chip. Now suppose that Peter does not change his mind, and meets

[22] It should be noted that one might find that compatibilism is true, *even in* a causally determined universe, because she feels that *we could have done otherwise*. In other words, she will say that even in a causally determined universe, we have the ability to have done otherwise in a very relevant sense. I return to this in the addendum which follows the paper.

Harry in the park, where they steal a bike. It certainly seems as though Peter is responsible for his part in the theft, since the desire to steal was his own (the chip had no influence over his decision), *but* he could not have done otherwise, since even if he had tried to, Harry would have activated the chip and *made* Peter steal the bike. Examples following this basic format have convinced many that alternate choices are not necessary for responsibility, and so encouraged the rejection PAP. Though we will leave PAP momentarily, it is an important principle, and its relationship to Frankfurt examples is a crucial point to which I will return later.

We can now return to discussion of Transfer NR. In an attempt to identify a situation in which no one is even partly responsible for some state, no one is even partly responsible for that state leading to another, but where, contra Transfer NR, someone is at least partly responsible for the fact that the consequence (i.e. the new state) arises, Fischer and Ravizza develop a sort of Frankfurt example which they call "Erosion."[23] In this example, Betty intends to plant explosives in a mountain at time t1, setting off an explosion which will cause an avalanche, and bury a military base below at time t3. Unbeknownst to her, a glacier has been eroding, and if she fails to detonate the explosives at t1, the final bit of erosion will occur at t2 and cause an avalanche which will bury the base at t3 (i.e., nature, rather than a person, assumes the role of counterfactual intervener). According to them:

> "(1) The glacier is eroding and no one is, or ever has been, even partly morally responsible for the fact that it is eroding; and
>
> (2) if the glacier is eroding, then there is an avalanche that crushes the enemy base at t3, and no one is, or ever has been, even partly morally responsible for the this fact;
>
> But Given Betty's responsibility, it is *not* true that

[23] John Martin Fischer and Mark Ravizza, 157.

(3) there is an avalanche that crushes the enemy base at t3, and no one is, or ever has been, even partly morally responsible for this fact."[24]

While this is a counterexample of sorts, to my mind, these examples all have a terrible flaw, as they only expose a hole which van Inwagen will have little trouble patching. The reason is that Fischer and Ravizza seem to think that Betty was morally responsible for the base being buried, because that result obtained partly in virtue of her actions. She chose to hit her detonator at time t1, which caused an avalanche that would have occurred anyway. But in doing so, they are positing a state p1 for which someone *is* at least partly morally responsible, namely Betty, as she is certainly at least partly morally responsible for her choice to detonate the explosives. Further, it seems to be in virtue of this that we hold her to be responsible for the base being buried, as if she did not have some hand in the avalanche's occurrence (e.g. if the natural forces caused the avalanche one second before she was about to do it on her own) then we would not hold her responsible for the base being buried.[25]

Once we realize this, all that van Inwagen will have to do is change Transfer NR to read:

(1) N(p1, p2…pn)

(2) N(p1→q, p2→q…..pn→q)

(3) N(q)

In making the revisions shown above, he can change the principle to involve *all* prior states, while still maintaining the spirit of his argument, and the intuitive plausibility with it. In other words, the principle can now read "p1, p2…pn, and no one is even partly responsible for the fact that p1, p2…pn." Here, p2 could serve as the glacier's erosion,

[24] Ibid, 157
[25] Of course, we would still hold her responsible for pushing the detonator.

while p1 could serve as Betty's actions. This revision rules out counterexamples of the sort above, as Betty *is* responsible for some previous state; specifically p1. Just by making this simple change, van Inwagen is to be able to entirely avoid the threat of Fischer and Ravizza's counterexample.

Though these initial examples are at first compelling, I see no way to develop a successful counterexample. Any attempt would appear to require us to assume a compatibilist version of moral responsibility, which of course any incompatibilist worth his salt will note begs the question. With that being said, I find the search for counterexamples to Transfer NR to be unlikely to yield any interesting results. Not all is lost with regards to finding a response to Transfer NR, however, and we can turn to this now.

IV. Responsibility for Conditionals - Hope for Solving Transfer NR: Transfer NR is impressive in its simplicity and strength. Both premises appear beyond criticism, and so the question seems to generally revolve around whether or not the conclusion follows. While many have been led to pursue this route, I think that there might be another way, as one of the apparently obvious premises might *not* be correct when applied to a causally determined world. In this section I want to push the idea that we can be responsible for conditionals. Moreover, I want to suggest that we are able to exhibit this sort of responsibility *even in* a causally determined world, and that this can be shown without making any assumptions in favor of compatibilism. Of course, if this is true, then perhaps we can escape the force of Transfer NR, by showing its second premise to be inapplicable to our world, regardless of the argument form's validity.

Let us start with an example, which we can call "Kelsey the Car Thief." Kelsey roams High Street in Columbus, Ohio looking unassuming, but actually attentively and methodically picking out Ohio State students to free from their car keys. Now let us suppose that a student, Gabriel, is walking down High St., his keys slipped into the front pocket of his shorts. Further, let us imagine that Kelsey, being the experienced and astute car thief that she is, has constructed a powerful magnet above High Street's sidewalk, which without fail will yank the keys from any students unfortunate enough to pass underneath it. Gabriel walks down High Street on his way home. He passes under Kelsey's magnet, and loses his car keys to the powerful device.

Now it certainly seems to be the case that the conditional: "if Gabe walks down High Street, then he will lose his keys" is true (remember, the magnet does its job without fail). What is far less obvious is that no one is morally responsible for this. In fact, quite to the contrary, it seems that Kelsey is at least in part responsible for the truth of this conditional, as it was her choice to erect the magnet that (in part) made it true. Further, we can frame Kelsey's responsibility in a way that is neutral with respect to compatibilism and incompatibilism. We could say, for instance, that she made the choice to put up and disguise the magnet, and we could say that this was a choice that Kelsey was aware that she was making, under normal and acceptable conditions for making such a decision. Her choice was not the result of drunkenness, any sort of mental deficiencies, manias etc.[26] Kelsey is responsible for the conditional in virtue of the fact that it was her *knowing choice* which in some way made the conditional true.

So what does this have to do with van Inwagen's transfer principle? Well, if we can be responsible for conditionals, then it is not immediately clear that Transfer NR will

[26] Though she steals often, it is not the result of kleptomania, but just the desire to make an easy buck.

be telling against the semi-compatibilist position. After all, maybe we can be responsible for conditionals in a deterministic world. Even though no one is responsible for the fact that p is true (where p is, again, some initial state), deterministic laws conjoined with p could certainly *result in* a knowing choice[27] which leads to responsibility for the truth of a conditional of the form p→q. Of course, if someone can be responsible for the truth of p → q regardless of her responsibility for the truth of p, then we can deny that the second premise in Transfer NR is applicable to a deterministic world, or more specifically that it *must* be applicable. Though this does nothing to challenge Transfer NR's validity, it shows that a premise can be plausibly denied, and so Transfer NR on its own is not telling against the compatibilist position.[28] If this is the case, van Inwagen cannot use it to help establish the incompatibility of moral responsibility and determinism, unless some *other* principle or proof is used to show that despite the counterexample above, the second premise of Transfer NR cannot be denied.

Unfortunately for the compatibilist, in van Inwagen's argument Transfer NR is *not* working on its own. After all, in the Direct Argument, the third premise is established in part via another principle:

(1) □ (S & L. → T)	From the definition of determinism
(2) □ (S→(L→T))	From 1
(3) N(S→(L → T))	*From 2 by way of A*
(4) NS	Assumption

[27] Since determinism, as noted above, does not rule out knowing choices.

[28] Incidentally, this is why I am not vulnerable here to the same criticism that I made of Fischer and Ravizza, namely positing a state for which someone is responsible as p. They could do this, were they to argue that p1 was a sort of knowing choice (e.g. Betty's igniting the explosives in "Erosion"), which was guaranteed to occur by the deterministic laws in conjunction with some previous state. However, in doing so, they would *not* be arguing against Transfer NR's validity, which is their aim, but rather against its applicability. The right approach is to do what I have done here, and merely suggest that deterministic laws can in some way introduce responsibility, and so challenge the applicability of Transfer NR to our world.

(5) N(L → T) From 3 and 4 by way of Transfer NR

(6) NL Premise

(7) NT From 5 and 6 by Transfer NR

Van Inwagen can admit that there are times when we can be responsible for conditionals. Nevertheless, if Transfer NR is valid (I have done nothing to show that it is not), then he can still proceed to show that we are *not* responsible for the truth of (L→ T) and so not responsible for the truth of T, *so long as Rule A* is acceptable. In other words, the incompatibilist still has the resources to attain the Direct Argument's conclusion. The reason is simple: Rule A establishes that no one is even partly morally responsible for the fact that "if S, then if L then T" (step 3). But if we assume NS, which seems right, then Transfer NR leaves us with N(L→T). Given that, the compatibilist in my position is left with nothing which will allow him to deny that no one is even partly responsible for the fact that L→T, and so NT follows. In short, my attempted denial of the applicability of the second premise above would be unhelpful, as Rule A will provide proof that Transfer NR's second premise *is* applicable to a causally determined world.

But the compatibilist is not without possible responses here. For one, he can attempt to deny Rule A, despite van Inwagen's assertion that its validity is "beyond dispute."[29] In attempting to illuminate the rule's certainty, he writes:

> No one is responsible for the fact that 49 x 18 = 882, for the fact that arithmetic is essentially incomplete, or, if Kripke is right about necessary truth, for the fact that the atomic number of gold is 79.[30]

One interesting thing to take note of here is that the three examples that van Inwagen uses all involve things which no one could be responsible for anyway, like the

[29] Peter van Inwagen, "The Incompatibility of Responsibility and Compatibilism," 32.
[30] Ibid, 32.

proposition “49 x 18 = 882.” I think part of the Rule A’s intuitive force comes from the fact that in contemplating its truth, we think of propositions which are true in all possible worlds, but are different from any which we encounter in debates about moral responsibility. In the next section, I aim to show that in fact Rule A is incorrect, and in doing so, cripple the Direct Argument.

V. Rule A – The Vulnerable Principle: The form for a successful counterexample to Rule A is straightforward: what is needed is a proposition which is true in all possible worlds, yet for which someone is at least partly morally responsible. From here, there seem to be two subtypes. The first is what I will call “multi-individually realizable.” For a multi-individually realizable proposition, the person (or group of people) who is responsible for its truth changes from possible world to possible world. It is true in all possible worlds, but there is no one person (or one group of people) responsible for its truth *in all worlds*. One such proposition is that stuff happens. This is true in all possible worlds, provided that we define stuff as “anything of the sort that it could happen,” or something along those lines. If it is true in all possible worlds that time passes and at that there is space, then *something* must happen in all possible worlds. It is not the case, however, that no one is responsible in any world for anything. After all, though it is true in all possible worlds that stuff happens, I am still partly responsible (if I can be responsible for anything) for the truth of the proposition that stuff happens, because I make knowing choices to bring stuff about. For instance, I am partly responsible for the truth of the proposition that stuff happens, in virtue of being responsible for writing this paper.

Notice here we have a case where a state obtains in all possible worlds, yet someone is still partly responsible for that state's obtaining. If writing this paper counts as doing stuff, and if I write the paper, then it certainly seems as if I make stuff happen. Further, if I am *responsible* for writing the paper, it seems that I am *responsible* for the truth of the proposition that stuff happens.

Though one might suggest that this example is flawed, as it is not true in all possible worlds that I write the paper, I can quickly point out that this is no problem for me whatsoever, as I only needed to find one proposition, that is true in all possible worlds (e.g. that stuff happens), for which someone is at least partly responsible. The fact that there is a different, more precise proposition (in this case, that a paper is written) which is not true in all possible worlds is not relevant. The proposition that stuff happens is still true in all possible worlds, and in writing the paper I am responsible for the truth of two propositions: (1) that stuff happens, and (2) that a paper is written. The presence of the latter does not in anyway affect the truth of the former.

One more serious worry here is how wide the responsibility must reach, and if propositions of the form "stuff happens" cast the net wide enough to constitute an attack on Rule A. Must I be responsible for p in every world, even in worlds which I do not exist? If van Inwagen means this by Rule A, there are two issues. First, it is not what he says, as Rule A holds as its conclusion that "p, and no human is even partly responsible for the fact that p." Second, it would then no longer do its job in the Direct Argument. "N" does not represent responsibility for a proposition's truth in all possible worlds. Therefore, van Inwagen's conclusion to Rule A would have to be something like Rp, rather than Np, where R means "p in all possible worlds, and no one is even partly

morally responsible in each and every possible world for the fact that p." The third step in the Direct Argument would then also need to be altered, to read: R(S→(L→T)) rather than N(S→(L→T)).

Were he to do this, however, he could no longer use the third step in conjunction with Transfer NR to establish N(L→T) or, two steps later in the Direct Argument, N(T). In other words, to plug the new step three into Transfer NR, he would have to make what we can call Transfer NRX, which would have to read Np, R(p→q) ⊢ N(q). The problem here is that the "stuff happens" example makes this look implausible. No one is responsible for the fact that a specific possible world exists. Further, no one is responsible in each and every world for the fact that it is true that, if the world exists, then stuff happens within it. Notice that these two assumptions fit perfectly into the first two premises of Transfer NRX. That being established, it seems wrong to say that no one is even partly responsible for stuff happening, though this is what Transfer NRX would have to hold as its conclusion. Remember, the response we are addressing essentially admitted that we can be responsible for stuff happening in an individual world, though not for it happening in all possible worlds, but this admission will end up harming the Direct Argument, since the "N" in the third premise of the Direct Argument would have to be changed to an "R." Once we do this, Transfer NRX cannot help establish the conclusion of the Direct Argument, which is N(T), in light of the fact that this move seems invalid, and in fact the argument no longer runs through.

But even if we did decide to look for a proposition in which someone was responsible for something being true in all possible worlds, our search would prove fruitful. It is at this point that I will introduce the second subtype of counterexample,

which I will call “individually realizable.” This simply means that there is one person, or a fixed number of specific people, who are responsible for a proposition’s truth *in all possible worlds*. To illustrate, imagine Laura the sound technician. Laura builds voice amplifiers, which one can speak into in order to increase his or her speaking volume. The following conditional, which I will at times refer to this as the “L-conditional,” seems to be true in all possible worlds:

> If Harlen speaks into a fully functional version of a voice amplifier invented by Laura, then his voice will be louder than it otherwise would have been.

Because of the definition of “amplifier,” it seems fair to assume that to be fully functional, the volume that the machine emits must be louder than the speaker’s voice, so in worlds where Laura has not found a way to do this (or where she has, but the machine is simply broken), the machine is not fully functional. If sound requires a medium, then in a world with no medium there is no fully functioning machine. In such a world, however, the conditional is still true since the antecedent is false. Given this, the L-conditional does seem to be the kind of thing which would be true in all possible worlds.

What stands out, however, is that it seems as though Laura *is* responsible for the truth of this conditional, i.e. the truth of this proposition which is true in all possible worlds. After all, *because of the way that she chose to design and make the machine*, if someone speaks into one of her machines in the proper way when it is fully functioning, then his or her voice becomes louder than it would have been without the machine’s assistance. It is because of Laura that the machine does what it does in the world in which it exists, and it is because it does what it does that if Harlen speaks into the machine, he

will be louder than he otherwise would have been. Well if this is all true, it also seems true that Laura is *responsible* for this conditional's truth.

If I am right, what this as well as the "stuff happens" example shows is that human beings can be at least partly responsible for propositions which are true in all possible worlds. If this is true however, then Rule A is false, and step three of the Direct Argument does not follow from step two. If this can be established, then van Inwagen will not be able to show that N(L→T), and so will not be able to use Transfer NR, regardless of its validity, to establish that NT. In short, the entire argument falls to bits. Further, the Laura example succeeds by using the idea that we can be responsible for conditionals. Because of this, coupled with the fact that the L-conditional is immune from the scope argument levied against the "stuff happens" example above, I will focus mostly on the L-conditional.

VI. Some Objections: I sense at this point that intuitions may vary greatly. I suspect that many will find themselves to be either unsure of whether what I have said has any merit, or unsure of why exactly their intuitions lead them to believe that it does. Another camp will likely want to dismiss the preceding section altogether, and levy several arguments to show that what I have said must be wrong. In this section, I will aim to affect the latter group, by anticipating what I find to be the more obvious, important and threatening counterarguments, and providing responses to these. Following my responses, which will hopefully shift the dissenters into the former category, I will give some reasons for why we do, or might, find my argument against Rule A intuitively plausible. I hope that this will not only relieve some personal unrest, but will also provide further justification for finding my rejection of Rule A successful.

A facet of my argument that might bother many stems from the apparent oddity that Laura is partly responsible for the truth of the L-conditional in all possible worlds, and so is responsible for this conditional even in worlds in which she does not exist. This is an attribute of all individually realizable propositions that do not contain agents that exist in all possible worlds. My opponent could attempt a *reductio ad absurdum* here, and say that since this is entirely ludicrous, the example fails. Though this would not speak to the "stuff happens" example, related issues could perhaps be brought to bear there as well. Furthermore, the conditional involving Laura is the proposition I have been working with, and so I would like to show that it will indeed work.

Admittedly, the idea of a person being responsible for something in a world in which he or she does not exist is more than just a little bizarre. However, it may be prudent to note that the entire concept of possible worlds could be classified as a bit weird, and so perhaps this should not be unexpected. I am of the opinion that if we take a step back, and look at just how it is that we think about possible worlds, it will become clear that Laura having transworld responsibility is not so crazy after all.

Barring the truth of modal realism, possible worlds are theoretical entities that we posit to explain necessity and possibility.[31] In our story, Laura is responsible for doing something that brings about a truth in each and every one of these theoretical entities. It is because of her choice to build the machine in a certain way that it amplifies the speaker's volume, and it is because it amplifies the speaker's volume when fully functional that the conditional is true in *all* possible worlds. In other words, her knowing choice made it the case that the conditional is true in all possible worlds, and since worlds in which she does

[31] In the event that modal realism is correct, no harm will come to my argument. It will simply mean that the possible worlds are not just theoretical entities, but concrete ones. Everything I have to say about them will still apply.

not exist are part of this set, it should seem natural to assign responsibility to her for the conditional's truth in these worlds as well. It is not as if, in the worlds in which Laura does not exist we say "the Laura in that world is responsible." In fact, we *never* say "the Laura in that world is responsible" for anything; rather we say "Laura is responsible for X, in that world," or "in that world, Laura is responsible for X." This clarification is crucial: it is not the Laura in each and every world which is responsible for the truth, but rather Laura is responsible in each and every world for the truth of the L-Conditional. In this way, we can evade the problem at hand.

Another related challenge would be to say that, "well sure, this conditional is true in all possible worlds, but that has nothing to do with Laura. The presence of the term 'fully functioning voice amplifier' in the conditional is sufficient for its truth, and Laura's presence adds nothing." In other words, the conditional "If Harlen speaks into a fully functioning voice amplifier then his voice will be louder than it otherwise would have been" is true in all possible worlds. Laura brings nothing to the equation, and so we cannot assign any responsibility to her.

It is indisputable that this proposition is true in all possible worlds, and further that no one is responsible for this fact. That being said, it would be fallacious to say that because no one is responsible for that proposition's truth, no one is responsible for the truth of the L-conditional, which is a *different* proposition. To illustrate the inadequacy of the inference above, let us imagine that it is true in all possible worlds that humans die. Humans, by definition say, cannot be immortal. Well there is a possible world where OJ makes a knowing choice to murder Nicole. Certainly we would not say in these

circumstances that OJ is not responsible for Nicole's death in virtue of the fact that she would have died anyway.

Of course, one might respond that OJ is not responsible for the fact that she one day died, but *is* responsible for the fact that she died in this way, at this time, etc., and this is entirely accurate, however it also bodes well for my argument. It is true in all possible worlds that fully functioning voice amplifiers make people's voices louder than they otherwise would have been. That being said, Laura is still responsible, in virtue of her ingenuity, engineering skill, etc., for the fact that *her* fully functioning voice amplifier amplifies sound. Just as OJ was responsible for his part in Nicole's death, Laura is responsible for her part in making the machines do what they do in all possible worlds where she builds fully functioning machines. Because of this, it would seem, she is responsible for the truth of the conditional "if Harlen speaks into a fully functioning version of a voice amplifier invented by Laura, then his voice will be louder than it otherwise would have been." The fact that a similar conditional would have been sufficient for its own truth is irrelevant to the question of Laura's responsibility for the truth of the L-conditional. It is partly because of Laura that if Harlen speaks into one of her machines under the proper conditions, his voice is louder than it otherwise would have been, and because of this we see her fingerprints on the truth of this conditional.

One might respond again by saying that these fingerprints are nowhere to be found in the worlds in which Laura does not exist, so once again my argument fails. As noted earlier however, we do not need the fingerprints of the Laura *in each particular world*, we merely need Laura to leave evidence of her presence in making the conditional true in all possible worlds. So long as she does this, her failure to be present in a certain

world renders the conditional trivially true, by making the antecedent false. It is because of what she does in some of the worlds in which she does exist (specifically, those in which she designs voice amplifiers that are fully functional, etc.) that the conditional is true in the worlds in which she does not. Her ability to have had a hand in making the conditional true in all possible worlds in no way depends on her *existing* in all possible worlds.

Another objection might go like this: Laura is not responsible for the truth of the L-conditional, but simply for the fact that the machine amplifies sound when fully functional. Here, however, my response is to point out that of course she is responsible for the fact that her machine amplifies sound when fully functional, but it is because of this that she is at least partly responsible for the truth of the L-conditional. In other words, *because* she is responsible for the fact that the machine amplifies sound in all worlds in which the proper conditions are met, she is at least partly responsible for the fact that if Harlen speaks into it, then his voice will be louder than it would have been otherwise.

Another stab at defeating my argument might proceed by suggesting that the conditional at hand is not really true in all possible worlds, and so not a counterexample to Rule A. Though I am skeptical about this, I think the L-conditional's truth in all possible worlds, is no more in doubt than that of premise two in the Direct Argument [which is: $\Box (S \rightarrow (L \rightarrow T)]$. It seems right to say that in all possible worlds, given S and L, T will obtain. However, perhaps some possible worlds are such that they have finite creatures (endowed with these powers by God) that can violate the laws of nature. Of course, this depends on how we define "laws," just as whether or not the L-conditional is true in all possible worlds depends on the definitions of "fully functioning" and

"amplifier." Either both are true in all possible worlds, in which case my objection to Rule A stands, or both are not true in all possible worlds, in which case the Direct Argument is unsound, as Premise two would be false. Any concern over whether this definition-dependent truth is true in all possible worlds, could just as easily be turned around on van Inwagen, whose argument appears to hinge on the very same sort of thing. It is my inclination to say that the definitions of laws and of fully functioning voice amplifiers allow for both of these conditionals to be counted as true in all possible worlds. Regardless, this challenge must either fail, or bring the Direct Argument down in addition to mine, rendering it counterproductive to my opponent's purposes.

VII. Strengthening Our Intuitions, With Frankfurt's Help: It is at this point that I would like to address those who think that there is something to my argument, but are not quite sure why. Intuitions are helpful, but they are a much more respectable philosophical tool if we can provide *some* explanation for why they exist, and of why we have the feelings about certain propositions that we do. With this in mind, something worth addressing right off the bat is the close connection between Rule A and the Principle of Alternate Possibilities, or PAP. As noted earlier, PAP is the principle which states that in order to be responsible for some action, the agent must have been able to do otherwise. If Kelsey *could not have behaved any differently* from the way that she did when she stole Gabe's car keys, then according to PAP, Kelsey is not responsible. As discussed earlier, Frankfurt style examples, like Peter the Bike Thief,[32] cast serious doubt on the truth of PAP by purporting to show that the ability to have done otherwise is irrelevant for the attribution of responsibility. Of course, Frankfurt examples do not *prove* that PAP is wrong, but they do lead many people to believe that it is. Provided that these

[32] See Section III.

intuitions are strong, I think the inclination towards thinking of PAP as a false principle leaves those who accept what Frankfurt has to say close to embracing the reasons I have given for treating Rule A as false as well.

As briefly noted earlier, van Inwagen chooses red herrings in attempting to show that we cannot be responsible for propositions which are true in all possible worlds. He is right that no one is responsible for the fact that 49 x 18 = 882, but this has nothing to do with the fact that this proposition is true in all possible worlds. In fact, no one is responsible for the fact that this equation is true for the very same reason that no one is responsible for certain other propositions that are most definitely *not* true in all possible worlds, for instance, that Neptune is farther from the Sun than Jupiter. It is the lack of human agency, and *not* the fact that it is true in all possible worlds, that renders 49 x 18 = 882, and the rest of the propositions enumerated by van Inwagen, such that no one is responsible for their truth. Because of this, these examples do nothing to provide support for Rule A.

To digress briefly but necessarily, an interesting question for van Inwagen is what exactly grounds Rule A's assertion that no one can be responsible for propositions which are true in all possible worlds. He certainly will not want to say that no one is responsible for such propositions because there is no human agency involved, as this could be undermined simply by finding a single such proposition that *does* involve human action, such as the L-conditional. Lack of knowing choice or informed human decision cannot be the reason either, since in Laura's case, a knowing choice is made. In fact, it seems that the reason van Inwagen finds it to be beyond dispute that we cannot be responsible for things which are true in all possible worlds, is that *it would not have been possible for*

things to have been otherwise, i.e. PAP! 49 x 18 could not have equaled other than 882, and it simply had to be true that if Harlen spoke into one of Laura's fully functioning amplifiers under the right conditions, then his voice would be louder than it would have been otherwise. PAP in some form, by all indications, comprises Rule A's foundation.

Of course, PAP is being employed in a slightly different manner in its support of Rule A. With the Frankfurt examples, what is being investigated is whether we can still be responsible for an action even when we do the same action *in all close possible worlds*. In other words, when Frankfurt says "we could not have done otherwise than X," he does not seem concerned to say that there is *no* possible world in which we do not do X. In the Peter the Bike Thief case for instance, there is a possible world where Peter decides not to go through with the theft, but just at that moment a vicious, long beaked bird pierces Peter's skull with his beak, doing no damage but removing the chip. This prevents Harry's signal from being received, allowing Peter to take the high road, and not steal the bike. Because such a possible world exists, there is a sense in which Peter could have done differently, though Frankfurt seems content to focus on our world and the ones which are nearby. Since such an encounter with a bird would require a world a good bit removed from our own, Frankfurt examples ignore cases such as this.

What van Inwagen does with Rule A is expand the scope of PAP to encompass *all* possible worlds rather than just *close* possible worlds, since now the propositions in question are true in all possible worlds. At first glance his expansion of PAP's scope from local to global possibility allows him to avoid the Frankfurt examples, since these focus on possibility which exists on a more local scale.

First glances can be deceiving however, and now it is time to bring back into play the strong intuitions regarding Frankfurt examples. What we have done with the "stuff happens" example and the L-conditional, essentially, is to construct Frankfurt style arguments *that respond to the global scale*. While the truth of the global type of examples may seem a bit less intuitively obvious than the truth of the local examples, this is likely due to the fact that it is much harder to think of propositions which are true in all possible worlds, for which we are nevertheless morally responsible, if for no other reason than because most of the propositions that we think of as being true in all possible worlds, are like the examples that van Inwagen gives, and void of human agency. That being said, all we are doing with the L-conditional is constructing an expanded Frankfurt example, which says that Laura is partly responsible for this conditional's truth, *even though she could not have prevented this conditional from being true*. It has to be true, yet she is still at least partly responsible.

With this I hope to have provided some idea of how exactly the examples given in the Section V are made to work, and why we might view them as successful in providing grounds for resisting Rule A. Given that one puts some stock in the conclusion brought about by Frankfurt style examples, he or she is but a short distance away from accepting what I have termed "global Frankfurt examples," i.e. those like the case of Laura the sound technician, which involve *all* possible worlds. If one is moved to accept the Frankfurt examples as the local level, she has good reason to do the same at the global level.

VIII. Conclusion: Responsibility for conditionals gets us quite a bit, at least with respect to finding an answer to the challenge posed by van Inwagen's Direct Argument

for Incompatibilism. The argument is compelling if both Transfer NR and Rule A are accepted, and so its success hinges on the success of the two principles. With regard to the former, our ability to be responsible for conditionals gives us the resources to challenge the soundness, though not the validity of the argument form. As noted, this is not a trivial accomplishment, since Transfer NR is somewhat telling against the compatibilist position even in the absence of the Direct Argument. Because of this, the ability to water down the principle is important, and this is exactly what we can do by showing that even if no one is responsible for some past state, someone *can* still be responsible for the fact that said state coupled with the laws of nature leads to a future state (in virtue of the knowing choices made by the person in question), and because of this, be responsible for the future state. In other words, we can dispute the applicability of the second premise of Transfer NR, by saying that this is simply not true of our world; we *can* be responsible for the conditional "if the laws of nature obtain, I will do X." The Kelsey the Car Thief example was meant to show that there is good reason to believe this.

Of course, challenging the applicability of the second premise will only get us so far, since as it turns out, if Rule A is accepted, then the second premise of Transfer NR *must* be applicable to a causally determined world. Fortunately, responsibility for conditionals allows us to exert force on van Inwagen's argument here as well. Rule A held that if some proposition is true in all possible worlds, then no one is responsible for said proposition. Though I was able to refute this with the "stuff happens" example, which did not involve a conditional, the focus was with the conditional "if Harlen speaks into a fully functioning voice amplifier invented by Laura then his voice will be louder

than it otherwise would have been." This conditional appears to be true in all possible worlds, yet it also seems as if someone, namely Laura, is at least partly responsible for it being true. I proposed this and believe myself to have provided acceptable answers to many important counterarguments. If I am correct, then Rule A fails, and with it the Direct Argument collapses. Further, the failure of Rule A allows me the room to challenge Transfer NR as an independent principle, which means that not only is the threat from the Direct Argument negated, but the challenge posed by Transfer NR is severely weakened.

What all of this means, to return to the very beginning of the paper, is that we might still be afforded the opportunity to exhibit moral responsibility in a world that is causally determined. At the very least, this strong and often cited argument against compatibilism will not succeed in relegating it to the heap of failed philosophical theories. Causal determinism, which may very well be true, is not in conflict with our desire to hold individuals responsible for their actions, at least not for the reasons given in the Direct Argument. While surely more needs to be said on this issue, the fall of the Direct Argument is a major step on the way to making the compatibilist view a plausible and tenable one.

Addendum: It may be recalled that at the beginning of Section II, I made it clear that strictly speaking, the issue at hand was not compatibilism (the belief that causal determinism is compatible with humans possessing *free will*), but semi-compatibilism. Despite this caveat, if my argument from semi-compatibilist principles is effective in

defending the compatibility of moral responsibility and determinism from the Direct Argument, then the compatibilist view is defended as well. This is because the compatibilist will agree with all of the tenets of semi-compatibilism that I have utilized in my argument against the Direct Argument. Further, it is hard to see how one could accept that free will is compatible with determinism, *without* accepting that moral responsibility is. Because of this, the compatibilist will hold both views, and thus the failure of the Direct Argument is a necessary condition for the truth of compatibilism. Keeping free will out of the picture allowed us to proceed without adding unnecessary complication to an already complex issue, all the while allowing us to defend compatibilism by ensuring a condition necessary for its success.

A question, however, is where the compatibilist can go from here. In what way can free will be introduced into the picture? In this addendum, I hope to briefly introduce this issue, and raise some possibilities for how one might go about doing this. Before proceeding, a brief note is in order: this section will not be an argument for a particular view of free will, nor even for compatibilism. I aim solely to investigate how one might go about moving from semi-compatibilism, to straight up, old fashioned, compatibilism.

Recall that the idea of a knowing choice was introduced precisely because it was neutral with respect to compatibilism and incompatibilism. It appears to be a conception of choice that is equally plausible on either view, and it entails neither. One path that the compatibilist can take then is to find a conception of free will that utilizes the idea of a knowing choice. The debate would then turn to whether or not the idea of a knowing choice could in anyway be adopted to help bolster a compatibilist conception of free

will.[33] While both sides could acknowledge that the idea of a knowing choice is a plausible one, they would still disagree over whether a knowing choice can ever be free (or in some way contribute to a free choice) in a deterministic world.

One reasonable place to look for answers is once again in the work of one of the two giants in this field of inquiry, Harry Frankfurt (the other, of course, is van Inwagen). In his paper "Freedom of the Will and the Concept of a Person,"[34] Frankfurt introduces two concepts, namely willing freely, and freedom of the will. When one wills freely, one has a second-order desire that the first-order one has be his or her will, where the will is "the desire(s) by which (one) is motivated in some action he performs, or by which he will or would be motivated when or if he acts."[35] In short, when one wills freely, she has the will she wants. For instance, Kelsey would be willing freely in the car thief example if her first-order desire were something to the effect of: "I desire that I have the keys of at least one Ohio State student," AND, if she had the second-order desire: "I want that my desire to have the keys of at least one Ohio State student be my will." When one has a second-order desire for a first-order desire to be one's will, Frankfurt calls this a "second-order volition." When one's second-order volition is realized, the person is willing freely.

This is in contrast to his concept of freedom of the will. One has freedom of the will:

> "Only if he is free to have the will he wants. This means that, with regard to any of his first-order desires, he is free either to make that desire his will, or to make some other desire his will instead. Whatever his will, then, the will of the person whose will is free

[33] Note that this would in no way threaten its status as neutral with respect to incompatibilism and compatibilism. The incompatibilist could perhaps also adapt it to in some way to help define free will on her view.
[34] Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," in *Free Will*, ed. Gary Watson (New York: Oxford University Press, 1982), 81-95.
[35] Ibid, 84.

> could have been otherwise; he could have done otherwise than to constitute his will as he did."[36]

At first, this seems to fly in the face of what we have previously said about rejecting PAP, and this may have been the case, had Frankfurt not to gone on to write:

> "It is a vexed question just how 'he could have done otherwise' is to be understood in contexts such as this one. But although this question is important to the theory of freedom, it has no bearing on the theory of moral responsibility…"[37]

In other words, Frankfurt accepts something like the semi-compatibilist conclusion here, where freedom of the will is unimportant to responsibility. Of course, the stated purpose of this addendum was to address free will, not to bolster semi-compatibilist claims with appeals to authority, and so we can turn to the former now. As Frankfurt writes, it is a "vexed question" as to how exactly we should interpret the ability to have done otherwise. Obviously the compatibilist will not want to say that the ability to have done otherwise requires that alternate courses of action had been physically possible, since then he would in fact not be a compatibilist at all, since no such possibility exists in a determined world.

One idea may be to once again introduce the idea of conditionals, and in particular, counterfactuals.[38] For instance, if determinism is true, then Kelsey had to erect the magnet. That being said, if the laws of nature were tweaked, or if the initial state had been slightly different, then perhaps Kelsey would have done differently. Compatibilism is not fatalism, and so it seems right to hold that if the laws or the initial state were different, then Kelsey's action (her knowing choice) might have been as well. It is in this sense that Kelsey could have done differently. This ability is still irrelevant to moral

[36] Ibid, 94.
[37] Ibid, 94.
[38] Timmothy O'Connor. "Free Will." *Stanford Encyclopedia of Philosophy*. 2005. http://plato.stanford.edu/entries/freewill/ (Accesssed September 12, 2008)

responsibility by Frankfurt's lights, though it is most certainly relevant to her having freedom of the will. If Kelsey could have done differently, in the sense that if it had been the case that she had a second-order volition that a first-order desire different from the one she actually had be her will, then this different first-order desire would have been her will, then she would have freedom of the will.

This thought is inspired by David Lewis, who in his magnificent paper "Are We Free to Break the Laws?" (incidentally, a challenge to van Inwagen's "The Incompatibility of Free Will and Determinism") suggests that there are two ways which we can think about our ability to break laws. There is the weak way in which I can break a law, which goes "I am able to do something such that, if I did it, a law would be broken," as well as the strong way "I am able to break a law."[39] Since we are not talking about Kelsey's ability to break laws, the specifics of Lewis's paper are not important here. The point to take away from this, is that Kelsey's ability to have had a different second-order volition is *similar* to the weak thesis in Lewis's paper, in that she is able to have another second-order volition in such a way that, were she to have this second-order volition, some other first-order desire would have been her will.

If a compatibilist could effectively argue for this point, what he would essentially be doing is explaining a way by which a knowing choice could be a free one. Of course, he would not be obligated to hold that all knowing choices are free. Still, he could provide a formula by which *some* knowing choices can be free, including choices by the likes of this essay's villain, Kelsey. In short, Frankfurt's view allows us to accept the semi-compatibilist thesis (since freedom of the will on his view is not relevant to moral

[39] David Lewis. *Philosophical Papers Volume II.* (New York: Oxford University Press, 1986), http://ebooks.ohiolink.edu (accessed October 6, 2008).

responsibility), while leaving us room to find free will in the world. In fact, when the semi-compatibilist accepts what Frankfurt says, and goes on to find a way in which free will could exist in a deterministic world, he is not so much making a *jump* to compatibilism, as much as he is accepting additional beliefs which commit him to compatibilism.

Of course, what I have suggested about free will, and particularly the meaning of "could have done otherwise," is highly contentious. One could still believe in some form of PAP for instance, holding that one cannot be free unless one had the ability to do otherwise in a strong, physically possible sense. However, it is not even necessary for one to be an incompatibilist to find a deficiency in Frankfurt's view. In fact, a fellow compatibilist, Gary Watson, does just this in his paper "Free Agency."[40]

Watson sees the relationship between free agency and our desires in a much different way.[41] On his view the platitudinal or commonsense view of freedom, which says that a person is "free to the extent that he is able to do or get what he wants,"[42] is more or less correct given that it is suitably qualified. Further, he holds that it can be used to develop a compatibilist stance. Watson proceeds by outlining two systems. The first is the valuational system of an agent, defined as:

> "that set of considerations which, when combined with [an agent's] factual beliefs (and probability estimates), yields judgments of the form: the thing for me to do in these circumstances, all things considered is *A*."[43]

The second is the motivational system of an agent, defined as:

> "that set of considerations which move [an agent] to action."[44]

---

[40] Gary Watson, "Free Agency," in *Free Will*, ed. by Gary Watson (New York: Oxford University Press, 1982), 96-110.
[41] A full discussion of why and how he disagrees with Frankfurt is not necessary for our purposes. Watson discusses this in depth in Section III of "Free Agency."
[42] Ibid, 96.
[43] Ibid, 105.

For Watson, we exhibit free agency when our motivational system and our valuational system align, i.e. when one is motivated by the judgments made by his or her valuational system. To quote Watson once more:

> "The possibility of unfree action consists in the fact that an agent's valuational system and motivational system may not completely coincide. Those systems harmonize to the extent that what determines the agent's all-things-considered judgments also determines his actions."[45]

To illustrate with two extremes used by Watson himself, an unfree agent such as a kleptomaniac is unfree (in the right situations) in virtue of the fact that those desires and wants which motivate him seem to be entirely unrelated to that which he values. He might value honesty and be adamantly opposed to stealing. That being said, the kleptomaniac is motivated by something completely outside of these values, which compels him to steal anyway. In fact, even if he valued the skill and determination needed to steal a product, it would not be this value motivating him to act, but rather some compulsion that is entirely unrelated. The value and the desire would happen to coincide, but the desire would not be *for* the value.[46]

The other extreme would be an entirely free agent, namely God. For God (assuming that we are talking of the God of the Abrahamic religions) "the dependence of motivation upon evaluation is total."[47] There is absolutely no gap between these systems; he is motivated solely by what he values.

Notice that looking at freedom in this way is compatibilistic, as whether or not what we value and what we are motivated by align, is wholly independent of determinism. Moreover, once again we have a compatibilist account of free will that

---

[44] Ibid, 105-6.
[45] Ibid, 106.
[46] Ibid, 110.
[47] Ibid, 110.

seems to conjoin quite nicely with the idea of a knowing choice in general, and the example of Kelsey the car thief in particular. Let us suppose that Kelsey is an egoist, and that she thinks she would increase her own happiness/utility by erecting a magnet and taking the keys of Ohio State students. Further, perhaps the fact that she values what will increase her happiness/utility is what motivates Kelsey to act. In these circumstances, Kelsey is free, i.e., her knowing choice was a free one given these assumptions.

Again, on Watson's account the compatibilist can introduce free will without undoing or putting in harm's way any aspect of my argument. The knowing choice is able to do all that it must do, without dragging along any baggage that harms what we have said leading up to this addendum. Further, these positions do not require the compatibilist to give up any other aspect that the argument hinged on. Laura is still perfectly able to be responsible for the truth of the L-conditional, although she might now have a free will as well. We can still make stuff happen, though perhaps now we can exercise free will in doing so.

In conclusion, the path from the semi-compatibilist view to the compatibilist view need not be a long one. The Direct Argument for Incompatibilism *must* be wrong, if compatibilism, semi- or not, is to be true. With that being said, provided that compatibilism does not entail anything that undermines my semi-compatibilist defense against the Direct Argument - and as I have hoped to show, it does not - then the fall of the Direct Argument clears the road for the jump from semi-compatibilism, to its more ambitious parent view. It is because of this that the main portion of my paper concluded by noting that semi-compatibilism's success against the Direct Argument makes the compatibilist view more tenable. In clearing an objection to the compatibility of moral

responsibility and determinism, it affords us the ability to more confidently look for the way in which free will and determinism are compatible, if at all. The job for the compatibilist now, is to find a theory of free will which is compatible both with determinism, and with our rejection of Rule A. In my mind, the prospects are bright.

Sources

Ayer, A.J., "Freedom and Necessity." In *Free Will*, edited by Gary Watson, 15-23. New York: Oxford University Press, 1982.

Chisholm, Robert M. "Human Freedom and the Self." In *Free Will*, edited by Gary Watson, 24-35. New York: Oxford University Press, 1982.

Fischer, John Martin and Ravizza, Mark. *Responsibility and Control*. New York: Cambridge University Press, 2000.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy v. 66*, no. 23. (December 4, 1969): 829-839.

Frankfurt, Harry G., "Freedom of the Will and the Concept of a Person." In *Free Will*, edited by Gary Watson, 81-95. New York: Oxford University Press, 1982.

Lewis, David. *Philosophical Papers Volume II*. New York: Oxford University Press, 1986. http://ebooks.ohiolink.edu (accessed October 6, 2008).

O'Connor, Timothy. "Free Will." *Stanford Encyclopedia of Philosophy*. 2005. http://plato.stanford.edu/entries/freewill/ (accessed September 12, 2008)

Ravizza, Mark, "Semi-Compatibilism and the Transfer of Non-Responsibility." *Philosophical Studies 75*, (1994): 62

Strawson, Peter, "Freedom and Resentment," In *Free Will*, edited by Gary Watson, 59-80. New York: Oxford University Press, 1982.

van Inwagen, Peter, *An Essay on Free Will*. New York: Oxford University Press, 1986.

van Inwagen, Peter, "The Incompatibility of Free Will and Determinism." In *Free Will*, edited by Gary Watson, 46-58. New York: Oxford University Press, 1982.

van Inwagen, Peter, “The Incompatibility of Responsibility and Determinism,” In *Action and Responsibility*, edited by Michael Bradie and Myles Brand, 30-36. Bowling Green Ohio: The Applied Philosophy Program at Bowling Green State University, 1980.

Watson, Gary, “Free Agency,” In *Free Will*, edited by Gary Watson, 96-110. New York: Oxford University Press, 1982. 96-110.